

***TOWARDS TEACHING EFFECTIVENESS  
ASSESSING STUDENT PERFORMANCE***

by

**Wm. Francis Herlehy III, Ph.D.,  
Associate Professor of Aviation Business Administration  
Embry-Riddle Aeronautical University  
College of Continuing Education, Wright-Patterson Center**

**Harold E. Gray, Ph.D., Editor  
Professor of Aviation Business Administration  
Embry-Riddle Aeronautical University  
College of Continuing Education, Offutt Center**



***TOWARDS TEACHING EFFECTIVENESS:  
ASSESSING STUDENT PERFORMANCE***

by

Wm. Francis Herlehy III, Ph.D.

Harold E. Gray, Ph.D., Editor

"A grade is an inadequate report of an inaccurate judgement by a biased and inconsistent judge of the extent to which a student has attained an undefined level of mastery of an unknown portion of an indefinite competence."

(Dressel, 1983)

**ABSTRACT**

Evaluation is a process that is used to measure the depth and breadth of specific knowledge and skills. It is used while the instructional process is taking place to indicate a degree of movement towards a desired performance. It is also used to indicate a level of achievement or degree of competence after the instructional process has taken place. Further, and most relevant to this presentation, is that it gives faculty an indication of the efficacy of their teaching methods.

The intended purpose of the author was to review the literature on how it is that adults learn and how the assessment of student performance should take place. With these as a premise, a survey of 142 faculty members, both full time and adjunct, was conducted to determine how they view (1) the assessment process they personally use and (2) the grading expectations of their respective academic institutions. The surveyed faculty were from six different universities.

The results of the descriptive survey were compared and contrasted with expectations developed from the literature review. The conclusions considered to be most significant were: (1) the meaning faculty gave to the concept that student performance assessments were a reflection of their teaching methods and (2) the preponderance of faculty who hold a personal belief or perceive an institutional expectation, or both, that ratings given for summative evaluations should conform to a "normal curve" distribution.

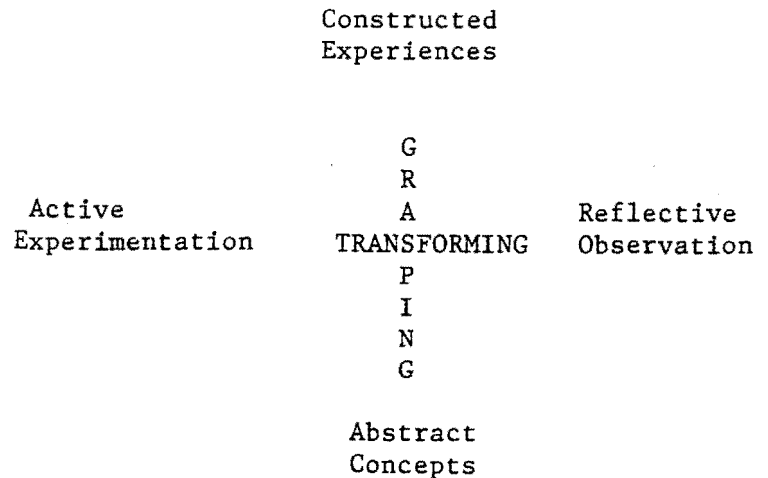
## HOW ADULTS LEARN

### Introduction

Before delving into the measurement of adults' learning performance, one needs to espouse some concept of how it is that adults learn as a premise for the ensuing discussion. Shockley's andragogical approach is but one of the accepted theories. Presenting it is not to disregard the others and not to ignore the issue of whether there is really any difference between how preadults and adults learn (pedagogy vs. andragogy: Knowles, et al). Shockley's model is but a point of departure.

The importance of espousing a specific learning theory is that it gives one a definitive description of the process the adult student is going through, thus, an infrastructure on which to build a particular instructional process. The collective achievement of some number of individual objectives is the goal of any instructional process. The related literature suggests that an assessment of student performance should be related to that goal and those objectives.

### SHOCKLEY'S MODEL



It is Shockley's belief that the adult learner comes to the non-traditional learning environment with certain "constructed experiences" (knowledge, skills, beliefs, etc.) and tends toward having more of these "experiences" and in more depth than does the traditional student.

Shockley presupposes the instructor of this person will have, or will quickly gain, an awareness of the students' "constructed experiences" and will begin the presentation of new "abstract concepts" by using the "experiences" as points of departure. In some other learning theories, a comparable notion is that the premise of this instructional process is to go from the "known to the unknown."

The adult learner, according to Shockley, will subject "abstract concepts" to either "reflective observation" or "active experimentation," or both. This is usually done with some assistance from the faculty member (the facilitator) and, perhaps, fellow students. The usual process would be to emphasize "reflective observation" when dealing with cognitive and/or affective changes and to emphasize "active experimentation" when dealing with what is primarily psychomotor change. They are not mutually exclusive.

The process of presenting "abstract concepts" using the "constructed experiences" as a point of departure and the subsequent "reflective observation" and/or "active experimentation" is recursive/reiterative until new "constructed experiences" are developed. The entire process is then repeated until all of the learning objectives have been subjected to the process.

During the process, formative evaluations can, and should, be used to determine if, and to what degree, progress or movement is being made on the transformation/change. At the end of the process, summative evaluations should be made to determine a level of achievement or degree of competence. How close is the observed behavior to the desired performance/ideal characteristic? Some of the other learning theories suggest different purposes.

## **ASSESSMENT OF PERFORMANCE**

As a lead in to a more in-depth discussion of performance assessment, we might ask a few rhetorical questions. When we assess students' performance are we attempting to find out:

What they know?

What they have learned?

How they learned (reacted to the process)?

How efficient/effective the instructor was?

Or some combination thereof.

### **Introduction**

College teachers are faced with the task of assessing the ability of students to recall specific knowledge or demonstrate certain skills and/or to use that knowledge or those skills in various specific ways (see Bloom's Taxonomy). The assessment of students' performance is broad and far-reaching. It has application in most aspects of university-level education and involves students in intellectual aspects well beyond memorization (rote). In this presentation, we will address both the assessment of students' performance in a somewhat natural, though not traditional setting, the adult learners' classroom, and then reporting on that assessment. Although this presentation and the examples used relate directly to the aviation-oriented, non-traditional, adult-learners' classroom, college teachers of almost all disciplines should find the ideas presented to be generalizable to their own settings and disciplines.

### **Competence**

Evaluation of students should seek to answer one question. "Is the individual competent?" Prior to answering this question, we must answer, "What is competence?" Saying that someone is competent and describing a level of competence is to imply there is a standard of behavior against which an observed behavior can be compared and defined (meets fully; adequate; or a letter grade with implied meaning). College teachers are responsible for judging students' degree of competency according to an agreed upon standard of performance. This is not an easy task.

### **Control**

The evaluation of an adult learner in the non-traditional classroom or, for that matter, in any other setting or for any other type of student is one of the most difficult tasks faced by the evaluator. Its difficulty lies in the attempt to simultaneously measure cognitive, affective, and psychomotor changes, which resulted from learning, and that are not necessarily mutually exclusive. There are difficulties in taking valid and reliable samples of student performance. There is difficulty in controlling the conditions under which the samples of performance are taken. Evaluation of student performance is one of the most complex areas of the teaching/learning process.

To help all those educators who assess student performance, four basic steps of performance evaluation will be discussed:

- a. Establish the purpose of the evaluation.
- b. Describe the standard's (ideal performance) characteristics.
- c. Measure/observe the actual student performance.

- d. Define/describe the level of achievement/performance.

### **Establish the Purpose:**

The first step in student performance assessment is to establish the purpose of the evaluation. In the reviewed literature, one finds both similar and divergent thought on the purpose of student performance assessment. The more widespread of these will be presented.

One commonly held view is that there are two types of student evaluations, each having a specific purpose. There are formative and summative evaluations. The formative evaluation should be conducted during the instructional process. Its purpose in assessment is to provide the student with on-going feedback which will improve his or her performance. There are also summative evaluations which are conducted at the end of a particular instructional process. Their purpose is to rate overall student performance. Frequently college faculty only think of this assessment in terms of assigning a letter grade at the end of a course of instruction. Properly done, it too could serve the purpose of improving student performance. Recognition that evaluations can provide students with on-going feedback and ratings at the end of a particular instructional process has many implications for teaching effectiveness and efficiency.

As previously mentioned, the purpose of formative evaluations of student performance is well-timed and on-going feedback. In general, feedback is most useful if given at the earliest opportunity after a specific performance. This feedback should be given openly and it should be descriptive rather than judgmental.

According to one study, correcting adult learners when they are wrong without belittling them was identified as a most helpful and critical teaching behavior.

Propriety does sometimes dictate that feedback be given away from other students. Regardless of when it is given, it should be specific and descriptive rather than judgmental. Furthermore, feedback should be directed at a behavior (performance) the student can do something about. Another rule of feedback is that it must be information that is understood by the student.

**NORM-REFERENCED SCHEMES FOR ASSESSMENT DO NOT HAVE A PLACE WITHIN FORMATIVE EVALUATIONS.** Motivational schemes, that do not objectively compare actual performance to a standard, are questionable. Criterion-referenced schemes provide a standard against which actual performance can be compared and, thus, specific feedback provided.

College faculty are familiar with summative (rating/ranking) evaluations. In the non-traditional setting, because the adult learner relies on the faculty for verifying/certifying their competence, summative evaluations can take on even more importance. Therefore, faculty must be committed to evaluations which are relevant, rigorous, and understandable and to conducting assessments that are reliable and valid and with ratings/rankings that are credible/believable to their students. No wonder the faculty feel the assessment of students is such an overwhelming task! It could be helpful to think of summative evaluations as a collection of evidence that allows faculty to have confidence in their judgement of student competency. The key

to this confidence is a multiplicity of objective data sources. The more objective and the more numerous the data sources, the more confident the faculty can be in the summative evaluation they make about the competence of students. Multiple objective data sources will be explored in this presentation under--"Measuring Student Performance."

### **Describe the Performance Standard's Characteristics**

In order to evaluate the performance of students, standards (the ideal characteristics) for the desired performance must be established. These ideal characteristics (standards) are best described as goals and objectives. Goals should serve as manifestations (specific written or spoken statements) of the ideal performance towards which students are moving. Objectives are statements of performance that provide evidence there is movement towards those goals. Objectives describe cognitive, psychomotor, and affective changes that can be either measured, observed, or demonstrated, which indicate movement towards the ideal character- of performance. Cognitive changes can be combined with psychomotor to become "proficiency" objectives. In this same way, "knowing" and "understanding" are interwoven with "applying" when students perform. Thus, both cognitive and psychomotor changes are assessed. In addition, affective changes can be interwoven and assessed.

Objectives constitute evidence of movement towards a goal. Thinking of objectives as evidence precludes thinking of them as ends. They are the means to an end, namely the goal.

Thus, after establishing the purposes of both formative and summative evaluations, we describe the ideal (desired) performance in terms of goals and objectives. The objectives must be both measurable and provide evidence of movement towards goals. Faculty know the objectives are comprehensive when attainment of all of them instills confidence the student has successfully reached the goals.

### **Measure/Observe Actual Student Performance**

Whether evaluating for improvement (formative) or for rating (summative), this step in the assessment process is the most difficult. Faculty are faced with the possible task of measuring cognitive, affective, and psychomotor changes simultaneously. The tools which best address this problem are those that provide concurrent measures. Some of these that will be succinctly presented are checklists, observation records, critical incident records, and anecdotal records. More information on each is available in the reviewed literature.



O  
B  
J  
E  
C  
T  
I  
V  
E  
S  
  
O  
T  
H  
E  
R  
  
A  
S  
P  
E  
C  
T  
S

SPECIFIC  
PREESTABLISHED

CHECK-  
LIST

CRITICAL-  
INCIDENT  
RECORD

NON-SPECIFIC  
NOT PREESTABLISHED

OBSERVATION  
RECORD

ANECDOTAL  
RECORD

### **Checklist:**

A checklist is the breakdown of a desired performance (change in behavior) into specific steps or more easily measured segments. The usual purpose is to standardize assessments, make observations easier, and documentation less time consuming. Frequently, the seminar/discussion group leader criticizes the checklist as not suitable for their particular teaching technique. Users of the technique who have used the checklist successfully have used it to keep record of student input, both quantity and quality, and to record demonstrations of the application of, and analyses using specific knowledge and skills. It provides evidence when doing either formative or summative evaluations.

### **Observations Record:**

The primary difference between an observation record and a checklist is nothing more than the observation record being evidence of unpredicted and non-specific, but yet relevant, input from students. Again, we are talking about maintaining a record of student competence for future evaluative purposes.

### **Critical Incident Records:**

Critical incident records can be used unto themselves or as part of the previously mentioned checklist. They would not be part of an observation record in that they are evidence of specific, and critical, learned behaviors. The critical incident record is used for cognitive, psychomotor, and affective change(s) that is(are) absolutely essential to being able to certify learning has taken place. These specific behaviors would likely be the difference between a passing and a failing grade.

### **Anecdotal Records:**

Anecdotal records are unlike the critical incident record as the checklist is unlike the observation record. The anecdotal record will provide evidence of unspecified, unpredicted, but yet relevant, behavior that indicates learning has taken place and/or a level of competence achieved.

The key to feeling confident about judgements of student performance is the use of multiple data sources. Each evaluation tool has both advantages and disadvantages. Faculty need to assess both independent and concurrent changes in cognitive, psychomotor, and affective behavior. They need to do so both during and after the instructional process. They need to make record of the assessment (assign a grade) for the student and other uses, not the least of which is a measure of instructor efficacy.

### **DEFINE/DESCRIBE THE LEVEL OF ACHIEVEMENT**

There are two prevalent methods by which college faculty assign student grades. These are "percentage grading" and "grading on a curve." There are, of course, other methods that are used by a relative minority of college teachers. McKeachie, 1986, one of the most respected assessment/evaluation authors, discusses several of these methods and the more relevant problems associated with them. In this presentation, we shall focus on the two more widely practiced systems with their numerous variants because most instructors fall neatly into the group that practices some variety of "percentage grading" or the group that, by one means or another, grades "on the curve."

Percentage grading systems are, perhaps, the most prevalent methods of assigning college grades. A faculty member announces some "absolute standard" early in a course, which takes the form of a percent of possible points that must be earned to obtain a specific grade. Advocates of such systems espouse the virtue of giving students advance notice of what is "expected of them" in order to earn certain grades. But does an announcement that a student "needs to obtain at least 90 percent to receive an 'A' in this course" really communicate what students need to learn? Does it define the domain of course content? Does it specify the difficulty of the test they will be given? No! to all of the above. This "advance information" concerning the "absolute standard" creates an illusion of informative clarity. It really tells next to nothing.

Most college-level course work falls into large, open, generalities-described content domains that do not lend themselves to meaningful interpretation of student performance using either raw or percent scores. Raw and percent scores are not only functions of content domain and student achievement, but are also artifacts of test difficulty. This size, openness, and lack of definitive description that characterizes most college level course content domains results in uncontrollable test item difficulty. Faculty are apt, whether intentionally or inadvertently, to develop a test on which no student is likely to attain at least a 70% or, just as likely, to develop one on which almost all students can attain at least 90%. Test difficulty is inherently norm-referenced, so to control for it we must violate the intrinsic nature of the content domains of most college-level course work, which is criterion

referenced. Yes, somewhat circuitous!

For the instructor who wants to use the raw or percent score there are some cases where establishing certain content domains will make these grades meaningful. "The domain description must be sufficiently detailed and delimiting to show clearly and definitively what facets of behavior are included, and more importantly, what behaviors are excluded from the domain" (American Psychological Association, 1985). "The fruitfulness of this orientation, content mastery, can only be realized when there is a possibility of defining the domain clearly and incisively so that the range of performance that lies within the domain can be definitively specified and agreed upon" (Thornhill, 1987).

Trying to make raw or percent scores fit content domains that cannot be described as suggested by the American Psychological Association or Thornhill will force one into arithmetical machinations or "adjustments," that are incompatible with the basic rationale of percent grading. They will be used by faculty to get themselves out of messes that a fundamentally illogical system got them into. It would have been better to avoid the predicament in the first place.

Like those who use "percent grading," instructors who grade on class curves usually value advance notice to students regarding what is required to receive one of various grades. They seek to do this with such statements as, "To receive an 'A', you must be among the top 15% of people in the class in total points amassed at the end of the term" ...at mid term"

Class-curve grading does not foster good interpersonal relations within a class. Having to "bump" others and being

"bumped" fosters ill will. Grading on a curve does not encourage group study or cooperative learning. It encourages isolation and exclusion. Curve grading does not motivate students to help one another learn. Quite the contrary, self interests would be 'best' served, in fact, by interfering with the learning of others in the group. A grading policy should not force students to compete with one another for grades. Learning is not inherently competitive. There is no logical reason why one student's success at learning must predispose others to less success or failure.

Other problems with class-curve grading are sample size and representativeness of the sample. One class in any particular course can be made up of many more "better" students than low achievers or the reverse might be true. A student in a class of more "low-achieving" students can all too easily rise to the top; more than a student can in a class with a majority of "better" students. Unless a class is truly representative, in the statistical sense, of the undergraduate or graduate population, to use curve grading is unfair and illogical.

With class sizes of 15 to 25 students, sampling error can be expected to make the difference of one letter grade for several students. Even in class groups as large as 50 students, sampling error can make a grade difference, though to a smaller number of students. The only way this sample-size concern can be abated is by evaluating several hundred students with the same measure (standardized tests). This creates an apparent paradox. The usual class size is too small a sample and lacks representativeness but yet curve grading is

espoused by some as the only logical foundation upon which to base grades for typical college courses (Hanna, 1984).

For those faculty who are going to use curve grading and do not have a large enough sample, the answer, it has been suggested, lies in the use of "anchoring." An "anchor" measure is a device with which a faculty member can judge or "take bearings" on the status of a particular class. To provide this "anchorage," a variable need only have the attribute that it correlate with performance in the course being graded. One example provided in the literature was the use of ACT/SAT/GRE scores (assuming they are available) as anchors because they are from a very large reference group and should correlate to performance in most college classes. Standardized tests are among others suggested. The large size of the reference group, it is suggested, will provide stability from sampling error and the correlation to performance will provide an adjustment to the statistical process used to obtain class curves for specific groups.

The concept of "anchoring" provides an answer to the problems of sample size, sampling error and lack of representativeness, but it does not remove the inherent competitiveness from the learning environment; it does not address predisposing one student to success and another to less success or, perhaps, failure; and it does not address the "preordination" of students' grades, regardless of student learning.

Some faculty seek to compromise the virtues and vices of the two major grading systems. Such approaches usually succeed in diluting the vices of each approach but at

the cost of also diluting the virtues of each.

The literature review gives rise to some criteria for judging college grading systems:

If grading is done on a curve, it will be subjected to scrutiny on:

The inherent competitiveness it brings to the learning environment.

The predisposition for success and failure it brings to the grading process.

The referencing/anchoring that is used to give relevance to the norm.

If grading is done on a class curve, it will be subjected to scrutiny on the sampling error attributable to sample size and the lack of representativeness to the sample.

Student cooperation in the learning environment should not be thwarted by a grading system that instills an artificial competition among peers, which, in fact, should be avoided. Grading cannot be a fixed-sum game.

Students should have a sense of control over their learning and over the grade that reports their achievement. They should know that a reported level of achievement reflects a certain degree of competence. The instructor should have a reciprocal sense of efficacy. (see note)

The grading system should be easily defined and interpreted. Its meaning should be communicable and consistent across as broad an academic spectrum as possible.

NOTE: To the instructor's "reciprocal sense of efficacy"

For the faculty member who strives to establish, maintain, and strengthen the relationships between subject matter, instructional objectives, assessments, and

reports of achievement, the grade attained by each and every student is an indicator of the degree of success to be enjoyed by the faculty member for achievement towards the optimum complementary relationship. The use of student grades, by a third party, to decide the efficacy of an instructor is questionable, at best. This is not to dismiss the harsh reality of the practice. It is an issue better addressed outside the cogent literature and within common practice.

#### **A DESCRIPTIVE SURVEY**

Based on ideas developed from the literature reviewed, a descriptive survey was developed to ascertain the attitudes, perceptions, and beliefs of selected faculty on the purposes of assessing student performance and the uses for the record of their achievement (grades).

A test bank of forty questions was developed and submitted to a panel of experts. This panel included four full time faculty with two different universities, four adjunct faculty with three different universities, and two faculty who retired from full time positions and now teach as adjuncts. Based on their comments and suggestions, a questionnaire was developed that consisted of twenty-five questions. The survey was administered to 38 faculty of the Air Force Institute of Technology for a pilot study. The intent of this study was to address potential process and/or content problems. The pilot study respondents were given the opportunity to make both written and verbal comments along with completing the survey itself. Their replies were not included in the study. However, their replies and comments were used to revise the questionnaire.

In its final form, two hundred copies of the questionnaire were provided to faculty from eight different universities. The copies, fifty of them, sent for distribution to full time and adjunct faculty of two specific universities were lost in the mail. The study was continued without giving them further consideration. Twenty-five copies of the questionnaire were distributed for faculty, both full time and adjunct, of six different universities. There were 142 acceptable questionnaires returned and included in the study.

FIGURE 1

SAMPLE

Dear Respondent,

I will be presenting a paper early next year on the role of student performance assessment in improving teaching effectiveness.

Your responses to the attached survey, if you choose to reply, will provide me with a faculty perspective on the subject, which will be an invaluable addition to the presentation.

I can assure you anonymity throughout the research and presentation to include not showing any specific relationships between academic institution and response.

My thanks in advance for your collegiality.

Wm. Francis Herlehy III, Ph.D.  
Associate Professor

P.S. I would be happy to provide you the collective results of the survey and my analysis if you care to have them. Include your request with the reply.





FIGURE 2

SAMPLE

Name (optional): \_\_\_\_\_

Faculty Status:      Full-time \_\_\_\_\_      Other \_\_\_\_\_  
                         Adjunct \_\_\_\_\_      (explain \_\_\_\_\_)

Academic Institution \_\_\_\_\_

(If more than one, please complete a separate survey for each or just complete one but only consider the institution listed when doing so.)

What do you consider your academic discipline?  
\_\_\_\_\_

How long have you taught at the college level?

Undergraduate \_\_\_\_\_ Graduate \_\_\_\_\_

What part of your teaching responsibilities are for:  
undergraduate classes? \_\_\_\_\_ graduate classes? \_\_\_\_\_ other? \_\_\_\_\_

What grading system do you use for purposes other than the grades you turn in at the end of the term?

i.e.      Letter grades \_\_\_\_\_  
            Letter grades w/plus or minus \_\_\_\_\_  
            Numerical grades (0-100) \_\_\_\_\_  
            Numerical grades (0-4) \_\_\_\_\_ How  
            many points after the decimal?  
            Other \_\_\_\_\_ Explain \_\_\_\_\_

On what basis do you assign these grades?

i.e.      To show student progress he/she has made \_\_\_\_\_  
            To reward student for achievement \_\_\_\_\_  
            To motivate student \_\_\_\_\_  
            To reinforce learning \_\_\_\_\_  
            To show what progress student has made relative \_\_\_\_\_

to the rest of the class \_\_\_\_\_

Other (explain) \_\_\_\_\_

What grading system do you use for the grades you submit at the end of the term?

i.e. Letter grades \_\_\_\_\_

Letter grades w/plus or minus \_\_\_\_\_

Numerical grades (0-100) \_\_\_\_\_

Numerical grades (0-4) \_\_\_\_\_ How many points  
after the decimal?

Other \_\_\_\_\_ Explain \_\_\_\_\_

On what basis do you assign these grades?

i.e. To show what the student knows \_\_\_\_\_

To show what the student has learned \_\_\_\_\_

To show how student compares to others in the class \_\_\_\_\_

To show level of competence \_\_\_\_\_

To conform to university grading standard \_\_\_\_\_

To show students contribution to the class \_\_\_\_\_

To motivate/reward student \_\_\_\_\_

Other (explain) \_\_\_\_\_

Does your university have a formal policy for grading?

Yes \_\_\_\_\_ No \_\_\_\_\_ How closely are you expected to follow  
that policy? \_\_\_\_\_

Do continuing enrollments in your classes play a conscious part in  
your grading system? Yes \_\_\_\_\_ No \_\_\_\_\_ Comment: \_\_\_\_\_

Do the grades you assign students indicate the effectiveness and/or  
efficiency of your teaching methods? Yes \_\_\_\_\_ No \_\_\_\_\_ If yes, how  
\_\_\_\_\_

How have you changed your teaching methods as a result of your  
students' grades? \_\_\_\_\_

In your opinion, how important are grades to your students?

Very important      Moderately important      Important      Not important

Why? \_\_\_\_\_

---

How important is the assessment of student performance, to include assigning grades, to you?

Very important      Moderately important      Important      Not important

\_\_\_\_\_

Why? \_\_\_\_\_

Any additional comment(s) on the survey subject: \_\_\_\_\_

\_\_\_\_\_

\_\_\_\_\_

Thanks for taking the time to respond.

Bill Herlehy



## DATA COLLECTION AND ANALYSIS

"Name" was included as an optional item on the questionnaire to ensure anonymity to those who desired it. Providing a "name" was taken as an approval from the respondent to be contacted for further comment and/or elaboration on responses to the survey. Primarily because of time constraints, a minimum amount of personal contact was made.

Of the 142 respondents, 56 were full time and 86 were adjunct. Seven of the faculty reported as adjuncts reported themselves as "other" but the explanation given warranted changing their category.

The academic institutions represented by the respondents were:

Embry-Riddle Aeronautical  
University  
Central Michigan University  
Park College  
Wright State University  
University of Dayton  
Xavier University

Coincidentally, this is somewhat of a cross-sectional representation of small to medium colleges and universities in the Midwest United States.

The academic disciplines represented were:

Architecture  
Accounting  
Marketing

Mathematics  
Psychology  
English Literature  
Communications  
Finance  
Philosophy  
Sociology  
Law  
Chemical Engineering  
Mechanical Engineering  
Aeronautical Engineering  
Organizational Behavior  
Human Resources Management  
Operations Management

Again, this is somewhat of a cross-sectional representation.

Teaching at the undergraduate level ranged from three to twenty-seven years with a mean of 10.7 years. All 142 of the respondents have experience teaching at the undergraduate level. Teaching at the graduate level ranged from five to twenty-two years with a mean of 8.2 years. Only 18 of the respondents had no teaching experience at the graduate level. The reason this question was included in the questionnaire was to indicate the respondents "qualifications" to make credible statements in regards to the subsequent questions.

Responses to the "part-of-teaching" question indicate that current teaching responsibilities range from solely undergraduate (30 respondents: 100%) to solely graduate (35 respondents: 100%). As worded, the question does not address whether the stated teaching responsibilities have anything to do with preference.

Re: "Grades during the term"  
"Letter grades" are used by 38 of the respondents and "letter grades w/plus or minus" are used by 56 of the respondents. "Numerical grades (0-100)" are used by 46 of the respondents but four of them explained that they were not using percentages but rather an achieved number of points against a possible number of points that was not 100 (i.e. 16 out of 20". No respondents reported using "numerical grades (0-4)" and no one reported using "points after the decimal." Two respondents reported using "other" and explained that they used descriptive comments as feedback to the student.

An assumption was made that formative grades are what is used in an arithmetical calculation to determine the summative grade. This assumption is based on the suggestion made by the panel of experts that faculty do not give comprehensive final exams but rather that they give a "final" formative exam and the grade for that assessment is somehow "summed" with the grade(s) for other formative assessments to determine the "final," or summative grade. The actual survey results seem to support this notion but there is an apparent paradox. If the assumption is credible, it is difficult to understand why faculty would use "letter grades," "letter grades w/plus or minus," or "descriptive comments" because there is no readily apparent or obviously meaningful way to "sum" these grades. This concern prompted personal contact with a small number of the respondents who gave a wide range of numbers they assign "in their head" to the different possible grades. Each of

them said it "was something that came from experience. Those contacted did indicate that the summative grade reported was derived by arithmetical manipulation of the formative grades, subjective as that system may be.

Re: "On what basis...assign grade"  
Even though they were not asked to do so, there was an expectation that respondents would select just one "basis" on which they assign grades for formative assessments. Seventy-two percent of the respondents selected at least two and 23 percent of them selected three.

"To show student progress..." was selected as a basis by 140 of the 142 respondents. This basis would certainly reflect the efficacy of the instructor. This is especially interesting because there is no indication that any of them gave a pretest. It seems what these faculty actually wanted to indicate was that their assessment is based on what the student knows or can do but not that it was necessarily learned as part of this particular instructional process. There is nothing inherently wrong with this. It does indicate faculty should rethink just how they consider this "basis" an indication of their effectiveness.

Only a small number of respondents selected the "achievement" (18) or "motivate" (12) basis for grades assigned to their formative evaluations. The other responses given for "basis" certainly indicate a majority view and they resemble each other. These two resemble each other but are quite unlike the other choices. They obviously represent a minority perspective.

However, I am not convinced this is not at least one basis for more faculty than those who did, in fact, select it. I feel that because using either of these as a basis for grades might not be a conscious act, it was not readily selected.

"To reinforce learning" was selected seventy-five times by the respondents. This is assumed to mean that the basis for the grade was to give the student an indication of the part, or parts, of the content domain they know and/or can use competently. We really do not know what knowledge and skills they brought into the instructional process.

For the thirty respondents who selected "...relative to the rest of the class" indications are that the basis on which they assign grades for formative evaluations is a class curve. Even those previously cited authors who are the staunchest supporters of grading on a curve do not support, in fact they recommend against, using a curve to grade on formative evaluations as being contrary to the basic premise of this type of evaluation.

Re: "Grading system...end of term"

One hundred and thirty-eight of the respondents indicated they submit "letter grades." Four indicated they submit "letter grades w/plus or minus" but went on to indicate that only the letter is used for the grade report. This was the expected response because usually the academic institution will prescribe the type of grade to be submitted and this is what almost all accredited colleges and universities prescribe. It is not a faculty preference.

Nobody opted for any of the other choices available.

Respondents indicated the basis on which they assign the end-of-term (summative) grade as follows:

To show what student knows .....	124
To show what student has learned.....	139
To show how student compares to others in class.....	100
To show level of competence.....	108
To show conformity with University standard.....	77
To show student contribution to class.....	62
To motivate/reward student.....	8
Other (explain).....	"part of the final grade is based on the effort put into the course by the student."

The basis "what student knows" is assumed to mean what part of the material covered in the course the student "knew" when assessed. It does not necessarily indicate what was learned as part of the instructional process because we do not know which knowledge/skills preexisted.

The basis "what student has learned," as previously discussed, can be an accurate basis for the summative grade if the faculty member has a measure of pre-existing knowledge and/or skills to be used for comparison at the end of the instructional process. Assuming that to be the case, grades assigned on that basis can also provide a valid

indication of the effectiveness of the instructor.

"How student compares to others in the class" is usually explained by faculty, who use this for a basis, to mean identifying which students fall where on the "normal curve" they are predisposed or preordained to fit. Some indicate they "adjust the curve" but none indicated any objective system to doing so (i.e. "anchoring" to compensate for sampling error). Again, "experience" was given as the best way for knowing how to adjust the curve.

"Level of competence" was an expected response. No explanations or comments were offered by those selecting this basis. Sixty-six percent of the respondents selected this as at least one basis for their end-of-term grades.

Fifty-four percent of the respondents selected "conformity to a University grading standard" as one basis for their summative grades. This was received with some ambivalence. None of them elaborated on what the standard/norm is. The small number of respondents who were personally contacted indicated their academic institution "expects a grade spread" and that they have developed an "anchor" through experience that seems to satisfy the institution. A common thought was that none of

the institutions seemed to check the "spread" very carefully--"a sense of direction without strong enforcement as one respondent put it.

Forty-four percent (62) of the respondents selected "contribution to the class" as one basis for their summative grades. Several of these respondents explained this response to mean "class participation." Of the respondents contacted, none had any means of recording "class participation" (i.e. checklist, anecdotal records, etc). They simply "remembered," Who participated. To what degree they participated. How cogent their participation was.

Each of the eight respondents who selected "motivate/reward" as a basis for their end-of-term grades went on to explain, in some fashion, that "the effort" a student put into the class played a distinct part in determining their final grade.

All respondents to the questionnaire indicated a "formal University standard for grading" that was used, at least, as a general guideline when assigning summative grades for their students. One-hundred and twenty of the respondents indicated they were expected to have a "grade spread."

To "continuing enrollments," forty of the respondents selected "yes"; eighty-two selected "no"; twenty selected neither and made no comment. Two of the respondents who selected "yes" commented that they were "expected to not be too tough."



All 142 of the respondents indicated the grades they assign show the effectiveness of their teaching methods. The respondents did not indicate so, but if I were to assume a cause and effect or causal comparative relationship between the effectiveness of teaching methods and summative grades, I would be foolish to expect any response other than what I got, especially on a self report. This result argues against final grades being a valid/reliable indicator of teaching effectiveness. What faculty member will assign a grade of 'C' or 'D' when doing so indicates a lack of effectiveness to their teaching methods? Let's not get ready to throw out the baby with the bath water! There is a way to use grades for this purpose. It will be discussed under "Conclusions."

Sixty respondents indicated on "changed your teaching" that they would spend more time covering material that an assessment indicated was a problem area for their students. Most of those indicated it would have to have been a problem area for "over some specified" percentage of the class. Only two-thirds of the respondents answered this question. All of the personal contacts stated they would adjust grades if there seemed to be a "problem area" pointed out by the assessment. One of the personal contacts stated she did not answer the question because she felt adjustments to her teaching methods should be made during the instructional process not at the end of it.

One hundred and eighteen of the respondents felt grades were "very important" to their students. Twenty-four of the respondents felt grades were moderately

important to their students. None of them opted for the other choices. The essence of the comments provided was that the importance of the grades related to the competitive and over-achieving nature of college level students.

All but six of the respondents indicated assessments and grading were "very important" to them. The six indicated they were moderately important. About half (68) of the respondents gave a reason for their selection. Most of the reasons given centered on a rather simplistic, "to let the students know how they did in the course." This and the comparable other reasons given are just too vague to suggest anything else.

## **CONCLUSIONS AND RECOMMENDATIONS**

Teaching effectiveness: the degree to which instructors can appropriately deal with individual student differences and instructional objectives in their teaching methods, wisely optimizing both the level of achievement and time it takes to master dimensions.  
(Hanna & Cashin, 1987)

Assessments are, to use a metaphor, suffering from a great illness. If the patient is to be cured he must admit his illness, so the "ills" of assessment cannot be cured until they are acknowledged. Until the "ills" are cured, the practitioner should question the validity of using assessments (1) to determine relative achievement in any specific course, (2) to determine the degree

of mastery of any competency, or (3) for determining the effectiveness of teaching methods. Of particular interest for this paper is "the effectiveness of teaching methods."

If there is anything that educational psychologists agree upon, it is that individual students differ. Effective teaching helps all students develop their talents to the maximum; it tends to increase their individual differences. In a given amount of time and with comparable effort, the more talented student will learn more than the less talented. EFFECTIVE TEACHING AT THE COLLEGE LEVEL CAN BE REFLECTED IN DIFFERING LEVELS OF ACHIEVEMENT BY DIFFERENT STUDENTS. With the major focus of this paper being on individual differences, attention will be given to providing a framework on which faculty can "appropriately deal with both the level-of-achievement dimension to assessments and the time-to-master-content dimension."

#### A Prescription:

**Subject Matter.** There are three types of subject matter content. (1) There is that which is completely specifiable, is masterable, and is essential. (B) There is that which is completely specifiable but either cannot or need not be mastered. (C) There is that which can neither be completely specified nor mastered. The first and third of these need to be translated into instructional objectives. The second type of subject matter could be included in a course content but does not warrant assessment. A first measure of teaching effectiveness is to

appropriately identify the subject matter as one of these types.

**Instructional Objectives.** As suggested by Mager (1975), instructional objectives should (1) state learning outcomes in terms of student behavior, (2) indicate the conditions under which the behavior should occur, and (3) specify the standard or level of achievement a student must attain. In keeping with Mager's criteria, another measure of teaching effectiveness is how well the instructional objectives relate to subject matter content. (i.e. (a) stroking techniques used in tennis can be very specific, completely mastered, and are essential. This calls for a "time to master" dimension to the instructional objectives and for specific teaching methods (demonstration and performance) vs. the "mental game" which is not easily or tightly specified and seldom mastered. This calls for a "level of achievement" dimension to the instructional objectives and possible manifestations of cognitive changes).

**Tests.** Subject matter described above in (A) and used as an example in (a) should be assessed with a mastery test because the content domain is definitive and essential to learning. This type of subject matter usually calls for only one form of the test because the student will rehearse/practice until they have mastered the subject matter. If the teaching has been effective, the differences in assessment results are most likely attributable to the time dimension. For subject matter described above in (C) and used as an example in (b), content can neither be specified clearly nor completely mastered so assessment should be based on

a demonstrated level of achievement. Teaching methods, which are used for this type of content domain, should emphasize progress checks and feedback. A measure of teaching effectiveness is the degree of relationship established between the subject matter, the instructional objectives, and the tests used for assessing either the degree of mastery or the level of achievement.

**Interpretation of Test Scores.** When using "raw scores," the instructor who has effectively maintained appropriate relationships between the just-mentioned factors of the instructional process will have a credible indication of their teaching effectiveness. The only concern, not normally addressed in the college-level classroom, that would make this indication of teaching effectiveness questionable is the lack of a measure of pre-existing knowledge and skills to be used for comparisons. These pre-existing knowledge and skills are the "constructed experiences" suggested by Shockley. Thus, they are not only important to accurately determining teaching effectiveness but also to the entire instructional process. It is the point of departure. That concern aside, both mastery and discriminating tests will indicate teaching effectiveness when using raw or percent scores.

Derived scores are obtained by converting raw or percent scores in any one of several ways to permit comparison with others being graded.

Using the reported survey indications, it is usually a raw or percent score that is used for the formative

evaluation. By definition, and with the one exception cited, this grade indicates learning effectiveness; by implication, it indicates teaching effectiveness. The expressed purpose of the derived score, which is usually used to report on the summative assessment, is to make "adjustments" for purposes of comparing and ranking those being graded. Thus, derived grades are not the measure one would want to use when seeking to determine teaching effectiveness. As suggested by one respondent, concerns for teaching effectiveness should occur during the instructional process not at the end of it. Raw and percent scores used to report on formative evaluations will give an indication, at the appropriate time, of teaching effectiveness and, if necessary, a signal for change.

## References

- American Psychological Association. (1985). Standards for educational and psychological testing. Washington, DC.
- Anastasi, A. (1982) Psychological testing. New York: Macmillan.
- Bloom, B. S. (1976). Human characteristics and school learning. New York: McGraw-Hill.
- Cox, W. F. and Dunn, T. G. (1989). Mastery learning: a psychological trap? Educational Psychologist 24, 14-29.
- Cronbach, L. J. (1991). Comments on mastery learning and its implications for curriculum development. Boston: Little, Brown.
- Dressel, P. Grades: one more tilt at the windmill. Memphis: Memphis State University.
- Ebel, R. L. (1986). Essentials of educational measurement. Englewood Cliffs, NJ: Prentice-Hall.
- Frisbie, D. A. (1979). Assigning course grades. Urbana-Champaign: University of Illinois.
- Guba, E. G. (1981). Effective evaluation. San Francisco: Josey-Bass.
- Hanna, G. S. (1984). Coordinating instructional objectives, subject matter, tests, and score interpretation. Manhattan KS: Kansas State University.
- House, E. H. (1980). Evaluating with validity. Beverly Hills: Sage Publications.
- Mager, R. F. (1975). Preparing instructional objectives. Belmont, CA: Fearon Publishers.
- McKeachie, W. J. (1986). Teaching tips. Lexington, MA: Heath.
- Milton, O. (1986). Making sense of college grades. San Francisco: Josey-Bass.
- Patton, M. W. (1990). Qualitative evaluation methods. Beverly-Hills: Sage Publications.
- Thornhill, B. L. (1987). Measurement and evaluation in education. New York: Wiley.